

GenCore version 5.1.7

Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2006, 18:07:43 ; Search time 2527.47 Seconds
(without alignments)
703.434 Million cell updates/s

Title: US-09-669-187A-148
Perfect score: 38
Sequence: 1 ggggactttccgcgtgggactttccaggggactttcc 38

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 179606

```
Minimum DB seq length: 0
Maximum DB seq length: 50
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

```
Database : EST:*
1: gb_esl1:*
2: gb_esl2:*
3: gb_esl3:*
4: gb_htc:*
5: gb_esl4:*
6: gb_esl5:*
7: gb_esl6:*
8: gb_esl7:*
9: gb_gsa1:*
10: gb_gsa2:*
11: gb_gsa3:*
```

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	16.2	42.6	46	1	A1544460	A1544460 fb75b10.x
2	16	42.1	25	1	A878803	A878803 of86e04.s
3	16	42.1	30	10	AG198773	AG198773 Pan trogl
4	16	42.1	49	9	CC178684	CC178684 RST325 Ba
5	15.6	41.1	37	8	R75772	R75772 V121a11.r1
6	15.4	40.5	28	1	A1416657	A1416657 sa10b04.x
7	15.4	40.5	34	1	A1357237	A1357237 qx63a08.x
8	15.4	40.5	34	9	A2760788	A2760788 IM0554F14
9	15.4	40.5	50	10	CG800335	CG800335 1118010B1
10	15	39.5	42	10	CG776618	CG776618 1123002C0
11	14.6	38.4	46	1	A1917121	A1917121 ts52a12.x
12	14.6	38.4	40	50	AU102736	AU102736 AU102736
13	14.6	38.4	50	1	AU106243	AU106243 AU106243
14	14.4	37.9	38	10	CL878011	CL878011 abf23bb12.
15	14.4	37.9	42	9	B2288740	B2288740 SALK 0221
16	14.4	37.9	45	9	A2345648	A2345648 IM0080K32
17	14.4	37.9	46	1	AA504387	AA504387 aa59c05.r
18	14.4	37.9	50	1	AU106577	AU106577 AU106577
19	14.2	37.4	35	10	CZ442924	CZ442924 IB3G02.f
20	14.2	37.4	37	11	TA379B06Q	TA379B06Q T. brucei
21	14.2	37.4	40	1	A1039253	A1039253 ox33b01.s
22	14.2	37.4	40	1	AU104208	AU104208 AU104208

C 96	13	34.2	33	9	AZ424092	AZ424092	1M0203D03	169	12.6	33.2	50	1	AUI02494
C 97	13	34.2	34	1	AI590768	AI590768	tw25b08.x	170	12.6	33.2	50	1	AUI02495
C 98	13	34.2	35	10	AJ599818	AJ599818	Arabidops	171	12.6	33.2	50	1	AUI02497
C 99	13	34.2	38	CV9679	CV9679	ya7bf04.si	172	12.6	33.2	50	1	AUI02499	
C 100	13	34.2	40	8	T49679	T49679	Pmpcm.060	173	12.6	33.2	50	1	AUI02504
C 101	13	34.2	40	10	CZ918363	CZ918363	4021009B0	174	12.6	33.2	50	1	AUI02508
C 102	13	34.2	43	1	AI421857	AI421857	tf55e11.x	175	12.6	33.2	50	1	AUI02509
C 103	13	34.2	43	1	AI590595	AI590595	tw23c11.x	176	12.6	33.2	50	1	AUI02510
C 104	13	34.2	43	1	AI613476	AI613476	ty37e03.x	177	12.6	33.2	50	1	AUI03928
C 105	13	34.2	46	8	W47560	W47560	zc35a01.rl	178	12.6	33.2	50	1	AUI07888
C 106	13	34.2	47	9	BH610050	BH610050	KG00327-5	179	12.6	33.2	50	1	AUI07889
C 107	13	34.2	50	1	AUI05329	AUI05329	AUI053329	180	12.6	33.2	50	1	AUI08000
C 108	13	34.2	50	1	AUI05338	AUI05338	AUI05338	181	12.6	33.2	50	1	AUI08001
C 109	13	34.2	50	1	AUI05340	AUI05340	AUI05340	182	12.6	33.2	50	1	AUI08002
C 110	13	34.2	50	1	AUI05355	AUI05355	AUI05355	183	12.6	33.2	50	2	BG391762
C 111	13	34.2	50	1	AUI05358	AUI05358	AUI05358	184	12.6	33.2	50	9	AZ366252
C 112	13	34.2	50	1	AUI05378	AUI05378	AUI05378	185	12.6	33.2	50	10	AG227817
C 113	13	34.2	50	1	AUI05735	AUI05735	AUI05735	186	12.6	33.2	50	11	CR043223
C 114	13	34.2	50	1	AUI05992	AUI05992	AUI05992	187	12.4	32.6	22	1	AA9311331
C 115	13	34.2	50	1	AUI06832	AUI06832	AUI06832	188	12.4	32.6	25	9	BZ378205
C 116	12.8	33.7	27	9	AZ473371	AZ473371	1M0289A18	189	12.4	32.6	28	1	AA152566
C 117	12.8	33.7	33	9	AZ352257	AZ352257	1M0090G11	190	12.4	32.6	30	9	AZ788334
C 118	12.8	33.7	36	9	AZ466529	AZ466529	1M0277K18	191	12.4	32.6	30	10	CZ472881
C 119	12.8	33.7	37	1	AI601822	AI601822	fc12d01.x	192	12.4	32.6	34	10	CZ476686
C 120	12.8	33.7	37	9	AZ346218	AZ346218	1M0081D16	193	12.4	32.6	37	1	AA921165
C 121	12.8	33.7	37	11	CT011645	CT011645	KBrH119N0	194	12.4	32.6	37	1	AA499210
C 122	12.8	33.7	38	3	BJ081977	BJ081977	BJ081977	195	12.4	32.6	37	2	BF211603
C 123	12.8	33.7	40	1	AJ683751	AJ683751	AJ683751	196	12.4	32.6	37	2	BF211603
C 124	12.8	33.7	41	7	CK814754	CK814754	Ra8g8c544	197	12.4	32.6	38	10	AJ622374
C 125	12.8	33.7	42	10	CL520536	CL520536	SAK2E07.F	198	12.4	32.6	40	1	AI924099
C 126	12.8	33.7	46	1	AA948496	AA948496	on53b06.s	199	12.4	32.6	42	6	W87969
C 127	12.8	33.7	46	1	AI284041	AI284041	qt72d04.x	200	12.4	32.6	42	8	CA797449
C 128	12.8	33.7	46	1	AI787001	AI787001	uj54601.x	201	12.4	32.6	43	9	AZ838833
C 129	12.8	33.7	46	10	AG202640	AG202640	Pan.tciol	202	12.4	32.6	46	1	AA831953
C 130	12.8	33.7	46	10	AL946937	AL946937	Arabidops	203	12.4	32.6	46	1	AA869036
C 131	12.8	33.7	49	1	AA663894	AA663894	ae74c09.s	204	12.4	32.6	46	1	AI619473
C 132	12.8	33.7	49	1	AI005180	AI005180	ou60a07.x	205	12.4	32.6	46	1	AL042972
C 133	12.8	33.7	50	1	AUI04206	AUI04206	AUI04206	206	12.4	32.6	48	9	AZ810742
C 134	12.8	33.7	50	1	AUI04856	AUI04856	AUI04856	207	12.4	32.6	48	9	BE736219
C 135	12.8	33.7	50	1	AUI05372	AUI05372	AUI05372	208	12.4	32.6	49	7	CK812785
C 136	12.8	33.7	50	1	AUI05730	AUI05730	AUI05730	209	12.4	32.6	50	1	AUI04860
C 137	12.8	33.7	50	1	AUI07888	AUI07888	AUI07888	210	12.4	32.6	50	1	AUI05386
C 138	12.8	33.7	50	1	AUI07889	AUI07889	AUI07889	211	12.4	32.6	50	9	AZ790119
C 139	12.8	33.7	50	9	AZ654289	AZ654289	1M0528E22	212	12.4	32.6	50	9	AZ955523
C 140	12.6	33.2	23	9	AZ585259	AZ585259	1M0390O14	213	12.2	32.1	31	1	AI286975
C 141	12.6	33.2	33	10	CZ469762	CZ469762	co5149-5p	214	12.2	32.1	35	11	TA275003P
C 142	12.6	33.2	34	1	AI811024	AI811024	tr03a03.x	215	12.2	32.1	37	1	AI357445
C 143	12.6	33.2	37	1	AI597737	AI597737	tu91b01.x	216	12.2	32.1	37	1	AA275360
C 144	12.6	33.2	38	9	AZ826498	AZ826498	2M0102N19	217	12.2	32.1	37	9	AZ343423
C 145	12.6	33.2	40	1	AA810284	AA810284	od14h07.s	218	12.2	32.1	39	10	CG724099
C 146	12.6	33.2	40	1	AI020727	AI020727	ua98h09.f	219	12.2	32.1	40	1	AA428089
C 147	12.6	33.2	44	8	D38738	D38738	HUMCIET818	220	12.2	32.1	40	9	AZ806060
C 148	12.6	33.2	44	10	AG193478	AG193478	fan.tciol	221	12.2	32.1	40	9	AZ663086
C 149	12.6	33.2	45	9	AZ465901	AZ465901	1M0276H09	222	12.2	32.1	41	9	AZ663495
C 150	12.6	33.2	46	1	AI958214	AI958214	fc92f08.y	223	12.2	32.1	42	3	BJ079239
C 151	12.6	33.2	46	1	AJ691901	AJ691901	AJ691901	224	12.2	32.1	42	9	AZ807828
C 152	12.6	33.2	47	3	BI826451	BI826451	603076143	225	12.2	32.1	42	10	CZ472457
C 153	12.6	33.2	47	8	N62591	N62591	zal3h10.81	226	12.2	32.1	42	10	CZ472457
C 154	12.6	33.2	47	9	BH850536	BH850536	SALK.0714	227	12.2	32.1	42	3	BJ067044
C 155	12.6	33.2	48	10	BU126913	BU126913	Danio rer	228	12.2	32.1	45	8	T67260
C 156	12.6	33.2	49	1	AUI253477	AUI253477	AUI253477	229	12.2	32.1	45	9	AZ345712
C 157	12.6	33.2	49	3	BM516312	BM516312	kj68e06.y	230	12.2	32.1	45	9	BH851251
C 158	12.6	33.2	49	3	BM566930	BM566930	kj07h07.y	231	12.2	32.1	46	1	AI554503
C 159	12.6	33.2	50	1	AUI02486	AUI02486	AUI02486	232	12.2	32.1	46	9	AZ454102
C 160	12.6	33.2	50	1	AUI02489	AUI02489	AUI02489	233	12.2	32.1	46	10	CL640369
C 161	12.6	33.2	50	1	AUI02492	AUI02492	AUI02492	234	12.2	32.1	46	9	BQ593365
C 162	12.6	33.2	50	1	AUI02494	AUI02494	AUI02494	235	12.2	32.1	47	5	BQ593365
C 163	12.6	33.2	50	1	AUI02495	AUI02495	AUI02495	236	12.2	32.1	48	2	BG685213
C 164	12.6	33.2	50	1	AUI02497	AUI02497	AUI02497	237	12.2	32.1	48	1	AA792088
C 165	12.6	33.2	50	1	AUI02499	AUI02499	AUI02499	238	12.2	32.1	49	1	AA442185
C 166	12.6	33.2	50	1	AUI02504	AUI02504	AUI02504	239	12.2	32.1	50	1	AL792018
C 167	12.6	33.2	50	1	AUI02508	AUI02508	AUI02508	240	12.2	32.1	50	1	AUI04323
C 168	12.6	33.2	50	1	AUI02509	AUI02509	AUI02509	241	12.2	32.1	50	1	AUI05719

242	12.2	32.1	50	1	AU106350	AU106350
243	12.2	32.1	50	1	AU106889	AU106889
244	12.2	32.1	50	1	AU106925	AU106925
245	12.2	32.1	50	6	CA969709	CCLX06824
246	12.2	32.1	50	9	AZ918144	100600300
247	12.2	32.1	50	10	CZ470013	C05470B-3
248	12	31.6	22	1	AA985475	OQ96E09.s
249	12	31.6	25	1	AW246315	2821832.5
250	12	31.6	26	9	AZ333170	1M0062B07
251	12	31.6	26	9	AZ615102	1M0444G15
252	12	31.6	32	10	CX194681	PSI7048-N
253	12	31.6	33	10	CG731153	1119134F1
c 254	12	31.6	34	9	BH789672	SALK 0423
255	12	31.6	34	9	BZ764154	SALK 1240
256	12	31.6	34	9	BZ764155	SALK 1240
257	12	31.6	35	9	AZ867805	2M0178J13
c 258	12	31.6	36	11	TA203A08P	
c 259	12	31.6	37	1	AI446151	Tj07ell.x
c 260	12	31.6	37	9	BH902529	SALK 0919
261	12	31.6	38	9	AZ305231	1M0005M15
262	12	31.6	38	9	AZ782406	2M0022N12
263	12	31.6	38	10	AJ597895	Arabidops
264	12	31.6	39	9	AZ626090	1M0466G12
c 265	12	31.6	40	10	AG205568	
c 266	12	31.6	40	8	H85898	YS92h07.s1
267	12	31.6	41	1	AV955327	AV955327
268	12	31.6	42	2	BF101499	601753311
c 269	12	31.6	42	6	CF643224	D61 D09 F
270	12	31.6	44	1	AV832733	AV832733
271	12	31.6	44	3	BJ001541	BJ001541
272	12	31.6	45	1	AL043028	DKFZp434P
c 273	12	31.6	45	10	CL640579	G077B12 G
274	12	31.6	46	1	AA206549	2G57F01.s
c 275	12	31.6	46	1	AA291212	Z618b12.8
276	12	31.6	46	9	AZ313975	1M0030K06
277	12	31.6	47	2	BES36289	601062681
278	12	31.6	47	10	AL760555	Arabidops
279	12	31.6	48	10	CL844507	EY12526-5
280	12	31.6	49	1	AA922626	om50a02.s
c 281	12	31.6	49	1	AI420261	tf06b06.x
c 282	12	31.6	49	2	BE960341	601653061
c 283	12	31.6	49	11	DMES46181	
284	12	31.6	50	1	AU102933	AU102933
285	12	31.6	50	1	AU104857	AU104857
286	12	31.6	50	1	AU104946	AU104946
287	12	31.6	50	1	AU105321	AU105321
288	12	31.6	50	1	AU105323	AU105323
289	12	31.6	50	1	AU105332	AU105332
290	12	31.6	50	1	AU105347	AU105347
291	12	31.6	50	1	AU105353	AU105353
292	12	31.6	50	1	AU105359	AU105359
293	12	31.6	50	1	AU105360	AU105360
294	12	31.6	50	1	AU105364	AU105364
295	12	31.6	50	1	AU105365	AU105365
296	12	31.6	50	1	AU105374	AU105374
297	12	31.6	50	1	AU105381	AU105381
298	12	31.6	50	1	AU105392	AU105392
c 299	12	31.6	50	1	AU106152	AU106152
c 300	12	31.6	50	1	AU106191	AU106191

ALIGNMENTS

RESULT 1
AI544460
LOCUS
DEFINITION
AI544460
IMAGE:3717691.3, similar to SW:CA19 HUMAN P20849 COLLAGEN ALPHA
1(IX) CHAIN PRECURSOR. ; mRNA sequence.
ACCESSION
VERSION
KEYWORDS

AI544460
fb75b10.x1 Zebrafish WashU MPMG EST
1(IX) CHAIN PRECURSOR. ; mRNA sequence.
AI544460.1 GI:4461833
EST.

46 bp mRNA linear EST 07-JUN-2001
Zebrafish WashU MPMG EST
Danio rerio cDNA clone
similar to SW:CA19 HUMAN P20849 COLLAGEN ALPHA
1(IX) CHAIN PRECURSOR. ; mRNA sequence.

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Danio rerio (zebrafish)
Danio rerio
Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 46)
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
CDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
ResourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Seq primer: T7 ET from Amersham
High quality sequence stop: 1
POLYA=No.
Location/Qualifiers
1..46
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:3717691"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLI-blue MRF"
/clone_lib="Zebrafish WashU MPMG EST"
/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
(5'-GCACCTAGTCTAGATCGGAGCGGCCCTTTTCTTTT3');
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."

Query Match 42.6%; Score 16.2; DB 1; Length 46;
Best Local Similarity 72.4%; Fred. NO. 1.1e+05;
Matches 21; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 GGGGACTTCCGCTGGGACTTTCAGGG 29
DB 9 GGGGATTCCGTTCCGGGGGTTCGGGG 37

```

RESULT 2
AA878803/c
LOCUS
DEFINITION
  AA878803      25 bp    mRNA      linear      EST 14-APR-1998
  of86e04.s1 NCI CGAP Li5 Homo sapiens cDNA clone IMAGE:1437246 3'
  similar to SW:EPDM HUMAN P10161 SALIVARY PROLINE-RICH PROTEIN PO
  ;contains TARI.t3 TARI repetitive element ;, mRNA sequence.

ACCESSION
AA878803
VERSION
AA878803.1 GI:2987768
KEYWORDS
EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Hominiidae; Homo.
REFERENCE
  1 (bases 1 to 25)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs@mail.nih.gov
  unknown library type
  Trace considered overall poor quality
  Insert Length: 185 Std Error: 0.00
  Seq primer: -40m13 fwd. RT from Amersham
  High quality sequence stop: 1.
  Location/Qualifiers
    1..25
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:1437246"
      /tissue_type="hepatic adenoma"
      /lab_host="DH10B"
      /clone_lib="NCI CGAP Li5"
      /notes="NotI; Cloned unidirectionally. Primer: Oligo dt.
      Average insert size 0.8 kb."

FEATURES
  source
    1..25
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:1437246"
      /tissue_type="hepatic adenoma"
      /lab_host="DH10B"
      /clone_lib="NCI CGAP Li5"
      /notes="NotI; Cloned unidirectionally. Primer: Oligo dt.
      Average insert size 0.8 kb."

ORIGIN
  Query Match      42.1%; Score 16; DB 1; Length 25;
  Best Local Similarity 79.2%; Pred. No. 1.3e+05;
  Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 10 CCGCTGGGACTTTCACGGGGAC 33
      ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25 CGCGGGGGGTCTTCCGGGGGCAC 2

RESULT 3
AG198773/c
LOCUS
DEFINITION
  AG198773      30 bp    DNA      linear      GSS 06-MAR-2004
  Pan troglodytes DNA, clone: RP43-079M17.T7, genomic survey
  sequence.
ACCESSION
AG198773
VERSION
AG198773.1 GI:45230949
KEYWORDS
GSS.
SOURCE
  Pan troglodytes (chimpanzee)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Hominiidae; Pan.
REFERENCE
  1
  Park.H., Kim.Y., Kim.S., Han.Y., Woo.T., Park.K., Eun.C.J.,
  Hoon.S.T., Chu.M., Kim.H., Joo.S., Kim.C., Song.W. and Yoo.H.
  BAC end Sequences of Library RP-43
  Unpublished
  2 (bases 1 to 30)
  Park.H., Kim.Y., Kim.S., Han.Y., Woo.T., Park.K., Eun.C.J.,
  Hoon.S.T., Chu.M., Kim.H., Joo.S., Kim.C., Song.W. and Yoo.H.
  Direct Submission
  Submitted (07-JAN-2002) Hong-Seog Park, Korea Research Institute of
  Bioscience and Biotechnology (KRIBB), Genome Research Center (GRC);

```

```

52, Oun-dong, Yusong-gu, Daejeon 305-333, Korea
(E-mail:redstone@mail.kribb.re.kr, URL:http://phs.grc.kribb.re.kr/,
Tel:82-42-866-7181, Fax:82-42-860-4409)
Clones are derived from the chimpanzee BAC library RP-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: T7
LIBRARY
  Vector      : pBACe3.6
  R.Site 1    : ECORI
  R.Site 2    : ECORI.
  Location/Qualifiers
    1..30
      /organism="Pan troglodytes"
      /mol_type="genomic DNA"
      /db_xref="taxon:9598"
      /clone="RP43-079M17.T7"
      /sex="male"
      /cell_type="lymphocytes"
      /clone_lib="RP-43 Chimpanzee Male BAC Library"

FEATURES
  source
    1..30
      /organism="Pan troglodytes"
      /mol_type="genomic DNA"
      /db_xref="taxon:9598"
      /clone="RP43-079M17.T7"
      /sex="male"
      /cell_type="lymphocytes"
      /clone_lib="RP-43 Chimpanzee Male BAC Library"

ORIGIN
  Query Match      42.1%; Score 16; DB 10; Length 30;
  Best Local Similarity 79.2%; Pred. No. 1.3e+05;
  Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 6 CTTTCGGTGGGACTTTCACGGG 29
      ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29 CTTTCCTCGAGGCGCTTTCACGGG 6

RESULT 4
CC178684/c
LOCUS
DEFINITION
  CC178684      49 bp    mRNA      linear      GSS 02-MAY-2003
  RST325 BayGenomics Gene Trap Library pGT2MpfA Mus musculus CDNA,
  mRNA sequence.
ACCESSION
CC178684
VERSION
CC178684.1 GI:30317235
KEYWORDS
GSS.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 49)
  BayGenomics.
  http://baygenomics.ucsf.edu/
  Unpublished (2001)
  Contact: BayGenomics
  Bay Area Functional Genomics Consortium (BayGenomics)
  Email: info@baygenomics.ucsf.edu
  Sequence tag generated by 5' RACE of total RNA from gene trap ES
  cell line. ES cell lines harboring insertion mutation of target
  gene are available upon request from BayGenomics. Annotation
  information available from
  http://baygenomics.ucsf.edu/cgi-bin/BaySearch.py?OPTION=EXACT&TYPE=
  CELL_LINE&KEY=RST325
  CELL LINE&KEY=RST325
  Class: Gene Trap.
  Location/Qualifiers
    1..49
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="129 ola"
      /db_xref="taxon:10090"
      /sex="Male"
      /cell_type="Embryonic stem cell"
      /clone_lib="BayGenomics Gene Trap Library pGT2MpfA"
      /notes="Vector: pGT2MpfA"

FEATURES
  source
    1..49
      /organism="Mus musculus"
      /mol_type="mRNA"
      /strain="129 ola"
      /db_xref="taxon:10090"
      /sex="Male"
      /cell_type="Embryonic stem cell"
      /clone_lib="BayGenomics Gene Trap Library pGT2MpfA"
      /notes="Vector: pGT2MpfA"

ORIGIN
  Query Match      42.1%; Score 16; DB 9; Length 49;
  Best Local Similarity 79.2%; Pred. No. 1.4e+05;

```

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 GACTTTCCGCTGGGACTTTCCAG 27
 |||||
 Db 26 GACTTTCCGCTGGGCTGCACCG 3

RESULT 5
 R75772 37 bp mRNA linear EST 06-JUN-1995
 LOCUS Y121a11_r1 Soares breast 2NbHBet Homo sapiens cDNA clone
 DEFINITION IMAGE:I58876 5' similar to SP:A42445 A42445 FSP27=FAT-SPECIFIC GENE
 FSP27 - ;, mRNA sequence.

ACCESSION R75772
 VERSION R75772.1 GI:850454
 KEYWORDS EST.
 ORGANISM Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 37)
 AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
 Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
 Trevaaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
 Wilson,R.

TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilton RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

Insert Size: 716
 High quality sequence starts: 1
 High quality sequence stops: 1
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Trace considered overall poor quality
 Possible reversed clone: similarity on wrong strand
 Insert Length: 716 Std Error: 0.00
 Seq primer: M13RP1
 High quality sequence stop: 1.

FEATURES Location/Qualifiers
 1..37
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="GDB:572930"
 /db_xref="taxon:9606"
 /clone="IMAGE:I58876"
 /sex="Female"
 /dev_stage="adult"
 /lab_host="DH10B (ampicillin resistant)"
 /clone_lib="Soares breast 2NbHBet"
 /note="Organ: breast; Vector: pT73D (Pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTTACCAATCTGAAGTGGAGCGCGCCCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of a modified pT73 vector (Pharmacia).
 Library went through one round of normalization to a Cot =
 230. Library constructed by Bento Soares and M.Fatima
 Bonaldo."

ORIGIN

Query Match 41.1%; Score 15.6; DB 8; Length 37;
 Best Local Similarity 81.8%; Pred. No. 1.9e+05;
 Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY
 Db

RESULT 6
 AI416657
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE 1 (bases 1 to 28)
 AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Corvett,V.,
 Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
 Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
 Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
 Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
 McCann,R., Waterston,R. and Wilson,R.

TITLE Public Soybean EST Project
 JOURNAL Unpublished (1999)
 COMMENT Other ESTs: sal0b04.y1
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

When it has been determined, an EST from the other end of this
 clone is listed in the 'Other ESTs on clone' field. Other ESTs:
 sal0b04.y1 GENOME SYSTEMS CLONE ID: Gm-cl003-8 Trace considered
 overall poor quality Possible reversed clone: similarity on wrong
 strand This clone is available through: Biogenetic Services, 801
 32nd Ave. Brookings, SD 57006 USA (phone: 800 423 4163; email:
 info@biogeneticservices.com)
 Seq primer: -40UP from Gibco
 High quality sequence stop: 1

POLYA-No. Location/Qualifiers
 1..28
 /organism="Glycine max"
 /mol_type="mRNA"
 /cultivar="Williams"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl003-8"
 /tissue_type="immature cotyledon"
 /lab_host="XL10-Gold"
 /clone_lib="Gm-cl003"
 /note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2:
 XhoI; Immature Cotyledon cDNA. The mRNA was isolated and
 then Stratagene's cDNA Synthesis Kit (catalog #200401) was
 used to synthesize the cDNA. First-strand synthesis was
 performed with 5-methyl dCTP, hence the ligated cDNA is
 hemimethylated. Stratagene's first-strand synthesis primer
 was used (GAGAGAGAGAGAGAGAACTAGTCTCGAG(T)-18). After
 second-strand synthesis, the cDNA ends were 'polished'
 with clone Pfu DNA polymerase, ligated to EcoRI adaptors,
 and phosphorylated. The XhoI site within the first-strand
 synthesis primer was restricted by digestion with XhoI;
 all XhoI sites in the cDNA would be protected by their
 hemimethylated status. The cDNA constructs were
 size-fractionated with a 500bp cutoff, using GibcoBRL Life
 Technologies' cDNA Size Fractionation column. The column
 eluent was then ligated into Stratagene's pBluescript II
 XR Predigested vector (pBluescript II SK(+)) that had been

FEATURES source

digested with EcoRI and XhoI, and phosphorylated). Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. Blue colonies 9n-15) have been sequenced, and possess putative cDNA inserts. This library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr. Paul Keim) 520-523-1372 (Virginia H. Coryell) FAX: 520-523-7500 email: paul.keim@nau.edu and also Dr. Lila Vodkin from the virginia.coryell@nau.edu and also Dr. Lila Vodkin from the University of Illinois"

ORIGIN

Query Match 40.5%; Score 15.4; DB 1; Length 28;
Best Local Similarity 76.0%; Pred. No. 2.2e+05;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 TTTCCTGGGGGACTTTCCAGGGGG 31
||||| ||||| ||||| ||||| |||||
Db 2 TTTCGGGGGGCCCCCCCCAGGGGG 26

RESULT 7

AI357237/c

LOCUS

DEFINITION

qx63a08.x1 NCI CGAP GC4 Homo sapiens cDNA clone IMAGE:2006006 3', similar to SW:PRP2_HUMAN P02812 SALIVARY PROLINE-RICH PROTEIN PRECURSOR ; contains element MSRI repetitive element ;, mRNA sequence.

ACCESSION

AI357237

VERSION

AI357237.1 GI:4108858

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 34)

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>, National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

TITLE

Unpublished (1997)

JOURNAL

Contact: Robert Strausberg, Ph.D.

COMMENT

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.

cdNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality

Insert Length: 1585 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 1.

FEATURES

source

1. .34

/location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2006006"

/tissue_type="pooled germ cell tumors"

/lab_host="DH10B"

/clone_lib="NCI CGAP GC4"

/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by

ORIGIN

Query Match 40.5%; Score 15.4; DB 1; Length 34;
Best Local Similarity 66.7%; Pred. No. 2.2e+05;
Matches 22; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 6 CTTTCCTGGGGGACTTTCCAGGGGGACTTTCC 38
||||| ||||| ||||| ||||| |||||
Db 33 CTTCCCCCGGGGGGGCGGGGGGCTTTTCC 1

RESULT 8

AZ760788

LOCUS

DEFINITION

1M0554F14R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0554F14 R, genomic survey sequence.

ACCESSION

AZ760788

VERSION

AZ760788.1 GI:12869009

KEYWORDS

GSS.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 34)

AUTHORS

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0554 row: F column: 14

Seq primer: CACACAGGAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 34.

FEATURES

source

1. .34

/location/Qualifiers

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC1M0554F14"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"

/clone_lib="Mouse 10kb plasmid UUGC1M library"

/note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into

ORIGIN									
Query Match	40.5%;	Score	15.4;	DB	10;	Length	50;		
Best Local Similarity	66.%;	Pred. No.	2.3e+05;						
Matches	22;	Conservative	0;	Mismatches	11;	Indels	0;	Gaps	0;
QY	4	GACTTTCGCTGGGACTTTCCAGGGGACTTT	36						
Db	45	GTCTTTCGGTGGACAGCTTCGCTGCTACTTT	13						

RESULT 11	AI917121/c	AI917121	46 bp	mRNA	linear	EST 28-JUL-1999
LOCUS	t552a12.x1	NCI CGAP Kid8 Homo sapiens cDNA clone IMAGE:232190.3,				
DEFINITION	similar to SW:CAJ3	BOVIN P04258 COLLAGEN ALPHA 1 (III) CHAIN.				
		; contains element TARI repetitive element ;, mRNA sequence.				
ACCESSION	AI917121	AI917121.1	GI:5636976			
VERSION	EST.					
KEYWORDS	Homo sapiens (human)					
SOURCE	Homo sapiens					
ORGANISM	Organs					

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
JOURNAL Homidae; Homo.
COMMENT 1 (bases 1 to 46)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA library Preparation: Life Technologies, Inc.
cDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality

Seq primer: -40UP from Gibco

High quality sequence stop: 1.

FEATURES
source

1..46
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2232190"
/issue_type="renal cell tumor"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid8"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.2 Kb. Life Technologies catalog #:
11524-014"

ORIGIN

Query Match 38.4%; Score 14.6; DB 1; Length 46;
Best Local Similarity 69.0%; Pred. No. 4.7e+05;
Matches 20; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GGGGACTTTCGCTGGGACTTTCACGGG 29
||| ||||| ||||| ||||| |||||
Db 45 GGGCCCTTTCCTCCCGGGTTTTTTTGGG 17

RESULT 12

AU102736
LOCUS AU102736 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
DEFINITION CAS09420, mRNA sequence.
ACCESSION AU102736 50 bp mRNA linear EST 28-JAN-2004
VERSION AU102736.1 GI:13552257
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 50)
Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J.,
Hata,H., Ota,T., Isoqai,T., Tanaka,T., Morishita,S., Okubo,K.,
Sakaki,Y., Nakamura,Y., Suyama,A. and Sugano,S.
Diverse transcriptional initiation revealed by fine, large-scale
mapping of mRNA start sites
EMBO Rep. 2 (5), 388-393 (2001)

JOURNAL
PUBMED

COMMENT Contact: Yutaka Suzuki

Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: ysuzuki@ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and

Sugano,S. Construction and characterization of a full
length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2),
149-156 (1997).

FEATURES
source

Location/Qualifiers
1..50
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CAS09420"
/clone_lib="Sugano Homo sapiens cDNA library"

ORIGIN

Query Match 38.4%; Score 14.6; DB 1; Length 50;
Best Local Similarity 69.0%; Pred. No. 4.7e+05;
Matches 20; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GGGGACTTTCGCTGGGACTTTCACGGG 29
||| ||||| ||||| ||||| |||||
Db 15 GGGGAGGGTCTCTCGGGTCTCTGGCAGG 43

RESULT 13
AU106243/c

LOCUS AU106243 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
DEFINITION ADSE00166, mRNA sequence.
ACCESSION AU106243 50 bp mRNA linear EST 28-JAN-2004
VERSION AU106243.1 GI:13555764
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 50)
Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J.,
Hata,H., Ota,T., Isoqai,T., Tanaka,T., Morishita,S., Okubo,K.,
Sakaki,Y., Nakamura,Y., Suyama,A. and Sugano,S.
Diverse transcriptional initiation revealed by fine, large-scale
mapping of mRNA start sites
EMBO Rep. 2 (5), 388-393 (2001)

JOURNAL
PUBMED

COMMENT Contact: Yutaka Suzuki

Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: ysuzuki@ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and
Sugano,S. Construction and characterization of a full
length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2),
149-156 (1997).

FEATURES
source

Location/Qualifiers
1..50
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="ADSE00166"
/clone_lib="Sugano Homo sapiens cDNA library"

ORIGIN

Query Match 38.4%; Score 14.6; DB 1; Length 50;
Best Local Similarity 62.2%; Pred. No. 4.7e+05;
Matches 23; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 GGGGACTTTCGCTGGGACTTTCACGGGCGGACTTTC 37
||| ||||| ||||| ||||| |||||
Db 37 GGAGACTCTGAGTCCGGGGTGGTGGGCGGACTTTC 1

RESULT 14

CL878011/c
LOCUS CL878011 Soybean random, unfiltered genomic library Glycine max
DEFINITION abf23b12.x1 Soybean random, unfiltered genomic library Glycine max

Fax: 858 558 6379
Email: ecker@alk.edu
This is single pass sequence recovered from the left border of
TDNA. This sequence lies within 300 bases of the 5' end of
A5c17610.

```

Class: link target="http://www.ncbi.nlm.nih.gov/nuclink/100881221"
Location/Qualifiers
  1..42 /organism="Arabidopsis thaliana"
        /mol_type="genomic DNA"
        /ecotype="Col-0"
        /db_xref="taxon:3702"
        /clone="SAIK 022130.29.10.x"
FEATURES
  source

```

```

/ecotType="COL-0"
/db xref="taxon:3702"
/clone="SALK_022130.29.10.x"
/clone.lib="Arabidopsis thaliana TDNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/rdna\_protocols.html"

```

```
/note="PCR was performed on each of which contained
```

each of which contains one or more DNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at <http://genom.sasb.edu/tadna/protocola.html>.

37.9% Score 14.4.4: DB 9: Length 42:

Query Match	37.9%	Score 14.4;	DB 9;	Length 42;
Best Local Similarity	75.0%;	Pred. No. 5.5e+05;		
Matches 18;	Conservative 0;	Mismatches 6;	Indels 0;	Gaps 0;
QY	14	TGGGACCTTTCCAGGGGGACTTTC	37	
Db	28	TGGGAAATTTCCATGGCGTCTCTC	5	

	AZ345648	45 bp	DNA	linear	GSS 29-SEP-2000
RESULT 16	LOCUS	1M0080K22F	Mouse 10kb plasmid UUGC1M library	Mus musculus genomic clone UUGC1M0080K22 F.	genomic survey sequence.
	DEFINITION				

ACCESSION	AZ345648
VERSION	AZ345648.1
KEYWORDS	GI:10424885 GSS.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus

REFERENCE

1 (bases 1 to 45)

Sciurognathi; Muroldes; Murinae; Mus.

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

TITLE
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
JOURNAL
 Unpublished (2000)
COMMENT
 Contact: Robert B. Weiss

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00

Plate: 0080 row: K column: 22
Seq primer: CGTTGTAAACACGCCAGT
Class: plasmid ends
High quality sequence stop: 45.

```

FEATURES
Location/Qualifiers
1..45
/organism="Mus musculus"
/mol_type="genomic DNA"

```

```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="DUGC1M0080K22"
/sex="Male"

```



```

RESULT 19
CZ442924/c
LOCUS      IB3G02.fwd HIV-vector integration sites from well-expressed
DEFINITION proviruses in human Jurkat T cells Homo sapiens genomic clone
IB3G02.fwd, genomic survey sequence.
ACCESSION  CZ442924
VERSION     CZ442924.1  GI:62379025
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 35)
AUTHORS    Lewinski,M.K., Blsgrove,D., Shinn,P., Chen,H., Hoffmann,C.,
Hannenhalli,S., Verdin,E., Berry,C.C., Ecker,J.R. and Bushman,F.D.
TITLE      Genome-wide analysis of chromosomal features repressing human
JOURNAL    J. Virol. 79 (11), 6610-6619 (2005)
PUBMED     15890899
COMMENT    Contact: Bushman FD
            Department of Microbiology
            University of Pennsylvania School of Medicine
            402C Johnson Pavilion, 3610 Hamilton Walk, Philadelphia, PA
            19104-6076 USA
            Tel: 215 573 8732
            Fax: 215 573 4856
            Email: bushman@mail.med.upenn.edu
            Class: PCR with specific primers.
FEATURES   Location/Qualifiers
            1..35
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /clone="IB3G02.fwd"
             /cell_line="Jurkat"
             /contig_lib="HIV-vector integration sites from
             well-expressed proviruses in human Jurkat T cells"
             /note="Vector: LTR-Tat-IRES-GFP (pEV731); We have
             investigated regulatory sequences in noncoding human DNA
             that are associated with repression of an integrated human
             immunodeficiency virus type 1 (HIV-1) promoter. HIV-1
             integration results in the formation of precise and
             homogeneous junctions between viral and host DNA, but
             integration takes place at many locations. Thus, the
             variation in HIV-1 gene expression at different
             integration sites reports the activity of regulatory
             sequences at nearby chromosomal positions. Negative
             regulation of HIV transcription is of particular interest
             because of its association with maintaining HIV in a
             latent state in cells from infected patients. To identify
             chromosomal regulators of HIV transcription, we infected
             Jurkat T cells with an HIV-based vector transducing green
             fluorescent protein (GFP) and separated cells into
             populations containing well-expressed (GFP-positive) or
             poorly expressed (GFP-negative) proviruses. We then
             determined the chromosomal locations of the two classes by
             sequencing 971 junctions between viral and cellular DNA.
             Possible effects of endogenous cellular transcription were
             characterized by transcriptional profiling. Low-level GFP
             expression correlated with integration in (i) gene
             deserts, (ii) centromeric heterochromatin, and (iii) very
             highly expressed cellular genes. These data provide a
             genome-wide picture of chromosomal features that repress
             transcription and suggest models for transcriptional
             latency in cells from HIV-infected patients."
ORIGIN
Query Match      37.4%; Score 14.2; DB 10; Length 35;
Best Local Similarity 70.4%; Pred. No. 6.4e+05;
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

CZ442924
IB3G02.fwd HIV-vector integration sites from well-expressed
proviruses in human Jurkat T cells Homo sapiens genomic clone
IB3G02.fwd, genomic survey sequence.
ACCESSION  CZ442924
VERSION     CZ442924.1  GI:62379025
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 35)
AUTHORS    Lewinski,M.K., Blsgrove,D., Shinn,P., Chen,H., Hoffmann,C.,
Hannenhalli,S., Verdin,E., Berry,C.C., Ecker,J.R. and Bushman,F.D.
TITLE      Genome-wide analysis of chromosomal features repressing human
JOURNAL    J. Virol. 79 (11), 6610-6619 (2005)
PUBMED     15890899
COMMENT    Contact: Bushman FD
            Department of Microbiology
            University of Pennsylvania School of Medicine
            402C Johnson Pavilion, 3610 Hamilton Walk, Philadelphia, PA
            19104-6076 USA
            Tel: 215 573 8732
            Fax: 215 573 4856
            Email: bushman@mail.med.upenn.edu
            Class: PCR with specific primers.
FEATURES   Location/Qualifiers
            1..35
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /clone="IB3G02.fwd"
             /cell_line="Jurkat"
             /contig_lib="HIV-vector integration sites from
             well-expressed proviruses in human Jurkat T cells"
             /note="Vector: LTR-Tat-IRES-GFP (pEV731); We have
             investigated regulatory sequences in noncoding human DNA
             that are associated with repression of an integrated human
             immunodeficiency virus type 1 (HIV-1) promoter. HIV-1
             integration results in the formation of precise and
             homogeneous junctions between viral and host DNA, but
             integration takes place at many locations. Thus, the
             variation in HIV-1 gene expression at different
             integration sites reports the activity of regulatory
             sequences at nearby chromosomal positions. Negative
             regulation of HIV transcription is of particular interest
             because of its association with maintaining HIV in a
             latent state in cells from infected patients. To identify
             chromosomal regulators of HIV transcription, we infected
             Jurkat T cells with an HIV-based vector transducing green
             fluorescent protein (GFP) and separated cells into
             populations containing well-expressed (GFP-positive) or
             poorly expressed (GFP-negative) proviruses. We then
             determined the chromosomal locations of the two classes by
             sequencing 971 junctions between viral and cellular DNA.
             Possible effects of endogenous cellular transcription were
             characterized by transcriptional profiling. Low-level GFP
             expression correlated with integration in (i) gene
             deserts, (ii) centromeric heterochromatin, and (iii) very
             highly expressed cellular genes. These data provide a
             genome-wide picture of chromosomal features that repress
             transcription and suggest models for transcriptional
             latency in cells from HIV-infected patients."
ORIGIN
Query Match      37.4%; Score 14.2; DB 10; Length 35;
Best Local Similarity 70.4%; Pred. No. 6.4e+05;
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY      12  GCTGGGACTTTCAGGGGACTTTCC 38
DB      27  GGTGAGTACTCTGCAGGAGGACGTTAC 1

RESULT 20
TA379B06Q/c
LOCUS      TA379B06Q      37 bp      DNA      linear      GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 379B06, reverse sequence,
genomic survey sequence.
ACCESSION  AL497861
VERSION     AL497861.1  GI:11873583
KEYWORDS   GSS.
SOURCE     Trypanosoma brucei
ORGANISM   Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE  1 (bases 1 to 37)
AUTHORS    Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.E., Rajandream,M.A. and Barrell,B.G.
TITLE      Direct Submission
JOURNAL
COMMENT    Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/T_brucei/.
FEATURES   Location/Qualifiers
            1..37
             /organism="Trypanosoma brucei"
             /mol_type="genomic DNA"
             /strain="TREU927"
             /db_xref="taxon:5691"
             /clone="379B06"
ORIGIN
Query Match      37.4%; Score 14.2; DB 11; Length 37;
Best Local Similarity 70.4%; Pred. No. 6.4e+05;
Matches 19; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY      12  GCTGGGACTTTCAGGGGACTTTCC 38
DB      36  GGTGGGATTTTCATGTTGATTTCC 10

RESULT 21
AI039253/c
LOCUS      AI039253      40 bp      mRNA      linear      EST 28-AUG-1998
DEFINITION OX33b01.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone
IMAGE:1658089 3', similar to SW:PF11_FIG P51524 PROPHENIN-1
PRECURSOR ;, mRNA sequence.
ACCESSION  AI039253
VERSION     AI039253.1  GI:3278447
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE  1 (bases 1 to 40)
AUTHORS    NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.

```


COMMENT Clones are derived from the chimpanzee BAC library RP-43 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS

Sequencing: TJ
LIBRARY
Vector : pBACe3.6
R-site 1 : EcoRI
R-site 2 : EcoRI
Location/Qualifiers

FEATURES

source
1..38
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-065002.TJ"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RP-43 Chimpanzee Male BAC Library"

ORIGIN

Query Match 36.8%; Score 14; DB 10; Length 38;
Best Local Similarity 66.7%; Pred. No. 7.7e+05;
Matches 20; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 3 GGACTTTCGGCTGGGACTTTCACGGGGGA 32
|||||
Db 8 GGNATATTGGGGGAATATTCTAGGGGGA 37

RESULT 27
AZ595836 40 bp DNA linear GSS 13-DEC-2000
LOCUS
DEFINITION
1M0408008R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0408008 R, genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
1 (bases 1 to 40)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
Niederhausern,A. and Wright,D.,Weiss,R.

Mus whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL
COMMENT
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0408 row: 0 column: 08
Seq primer: CACACAGAAACGATGACG
Class: plasmid ends
High quality sequence stop: 40.

FEATURES

source
1..40
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0408008"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, P-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 36.8%; Score 14; DB 9; Length 40;
Best Local Similarity 66.7%; Pred. No. 7.8e+05;
Matches 20; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GGGGACTTTCGGCTGGGACTTTCACGGGG 30
|||||
Db 11 GGGGCTGTGCAGCTTGTTCCTTCAGTGG 40

RESULT 28

AZ664762 40 bp DNA linear GSS 14-DEC-2000
LOCUS
DEFINITION
IM0545F18F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0545F18 F, genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
1 (bases 1 to 40)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
Niederhausern,A. and Wright,D.,Weiss,R.

Mus whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL
COMMENT
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA

Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0545 row: P column: 18
Seq primer: CGTTGTAAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 40.

FEATURES

source
1..40
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0545F18"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, P-"

/clone lib="Mouse 10kb plasmid UUGC1M library"
 /notes="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male); was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adapted DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (G14732114 [gb]AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adapted mouse DNA was annealed to
 adapted vector DNA, and transformed into
 chemically-competent *E. coli* XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN
 Query Match 36.8%; Score 14; DB 9; Length 40;
 Best Local Similarity 66.7%; Pred. No. 7.8e+05;
 Matches 20; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 8 TTCCGCTGGGACTTTCACGGGGACTTTC 37
 ||||| ||||| ||||| ||||| |||||
 Db 11 TTCCACTCTGGACATTCCTTGCTTCCTTC 40

RESULT 29
 CZ487957
 LOCUS 40 bp DNA linear GSS 29-APR-2005
 DEFINITION f05395-3prime Exelixis piggyBac WH insertions Drosophila melanogaster genomic sequence recovered from 3' end of piggyBac, genomic survey sequence.

ACCESSION CZ487957
 VERSION
 KEYWORDS
 SOURCE GSS.
 ORGANISM Drosophila melanogaster (fruit fly)

Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 40)
 AUTHORS Thibault,S.T., Singer,M.A., Miyazaki,W.Y., Milash,B., Dompe,N.A., Singh,C.M., Buchholz,R., Densky,M., Fawcett,R., Francis-Lang,H.L., Ryner,L., Cheung,L.M., Chong,A., Erickson,C., Fisher,W.W., Greer,K., Hartouni,S.R., Howie,E., Jakkula,L., Joo,D., Killpack,K., Laufer,A., Mazzotta,J., Smith,R.D., Stevens,L.M., Stuber,C., Tan,L.R., Ventura,R., Woo,A., Zakrajsek,I., Zhao,L., Chen,F., Swimmer,C., Kopczyński,C., Duyk,G., Winberg,M.L. and Margolis,J.
 TITLE A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac
 JOURNAL Nat. Genet. 36 (3), 283-287 (2004)
 PUBLISHED 14981521

COMMENT Contact: Roger A Hoskins
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory
 Mailstop 64-121, One Cyclotron Road, Berkeley, CA 94720, USA
 Tel: 510 486 4015
 Fax: 510 486 6798
 Email: RHoskins@lbl.gov

Sequence recovery method was inverse PCR.
 Sequence orientation is forward strand relative to 5' end of piggyBac element.
 The piggyBac insertion position is 1 in the 40 bases. This insertion position refers to the first base of the 4 base TTTAA target recognition sequence.
 Class: transposon insertion site.

FEATURES
 Location/Qualifiers
 1..40

/organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /strain="isogenic w- strain"
 /db_xref="taxon:7227"
 /clone_lib="Exelixis piggyBac WH insertions"
 /note="Vector: piggyBac WH (GenBank accession number AY515148); An isogenic w- Drosophila melanogaster strain was mutagenized by remobilization of transposable elements. We remobilized the WH element using the constitutive alpha-1 tubulin:piggyBac transposase source. We remobilized the WH element from a single ammunition element on the Binsyncy balancer chromosome in dysgenic females. We outcrossed dysgenic virgin females in vials to the isogenic w- strain and selected new hops in the following generation. All lines were mapped to a chromosome by standard genetic methods, examined for homozygous viability, and used for recovery of flanking genomic sequence by inverse PCR."

ORIGIN
 Query Match 36.8%; Score 14; DB 10; Length 40;
 Best Local Similarity 66.7%; Pred. No. 7.8e+05;
 Matches 20; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 5 ACTTTCCGCTGGGACTTTCACGGGGGACT 34
 ||||| ||||| ||||| ||||| |||||
 Db 3 RAATTTTCTGTGTTTTCAGTGTGACT 32

RESULT 30
 CC200179
 LOCUS 41 bp mRNA linear GSS 09-MAY-2003
 DEFINITION XG205 BayGenomics Gene Trap Library pGTLxf Mus musculus cDNA, mRNA sequence.

ACCESSION CC200179
 VERSION
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.
 BayGenomics.
 http://baygenomics.ucsf.edu/
 Unpublished (2001)

CONTACT: BayGenomics
 Bay Area Functional Genomics Consortium (BayGenomics)
 Email: info@baygenomics.ucsf.edu
 Sequence tag generated by 5' RACE of total RNA from gene trap ES cell line. ES cell lines harboring insertion mutation of target gene are available upon request from BayGenomics. Annotation information available from
 http://baygenomics.ucsf.edu/cgi-bin/BaySearch.py?OPTION=EXACT&TYPE=CELL_LINE&KEY=XG205
 Class: Gene Trap.

FEATURES
 Location/Qualifiers
 1..41
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="129 Ola"
 /db_xref="taxon:10090"
 /sex="Male"
 /cell_type="Embryonic stem cell"
 /clone_lib="BayGenomics Gene Trap Library pGTLxf"
 /note="Vector: pGTLxf"

ORIGIN
 Query Match 36.8%; Score 14; DB 9; Length 41;
 Best Local Similarity 77.3%; Pred. No. 7.8e+05;
 Matches 17; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 GACTTTCCGCTGGGACTTTTC 25

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert length: 672 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.

FEATURES

Location/Qualifiers
1..49

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2011857"

/tissue_type="glioblastoma (pooled)"

/lab_host="DH10B"

/clone_lib="NCI CGAP Brn23"

/note="Organ: brain; Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGGAGCGCGCATATCTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaudo."

ORIGIN

Query Match 36.8%; Score 14; DB 1; Length 49;

Best Local Similarity 60.5%; Pred. No. 8e+05;

Matches 23; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 GGGGACTTCCGCTGGGACTTTCAGGGGACTTTC 38

Db 39 GGGTTTTTCCCTTTTCCCTTTTGGGGGTTTTC 2

RESULT 34

BE311372

LOCUS

DEFINITION BE311372 50 bp mRNA linear EST 26-OCT-2000 601057775T2 NCI CGAP_Mam6 Mus musculus cDNA clone IMAGE:3158010 3', mRNA sequence.

ACCESSION

BE311372

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 50)

NIH-MGC <http://mgc.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LAM7650 row: e column: 19

High quality sequence start: 3

High quality sequence stop: 50.

Location/Qualifiers

1..50

/organism="Mus musculus"

/mol_type="mRNA"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:3158010"

FEATURES

source

Query Match

Best Local Similarity 36.3%; Score 13.8; DB 1; Length 28;

Matches 18; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GGGGACTTTCGCTGGGACTTTC 25

/sex="female, virgin"

/tissue_type="infiltrating ductal carcinoma"

/dev_stage="5 months"

/lab_host="DH10B"

/clone_lib="NCI CGAP Mam6"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;

Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Library constructed by Life Technologies. Investigator

providing samples: Jeffrey Green, M.D., NIH"

ORIGIN

Query Match 36.8%; Score 14; DB 2; Length 50;

Best Local Similarity 77.3%; Pred. No. 8e+05;

Matches 17; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8 TTCCGCTGGGACTTTCACGGG 29

Db 4 TACCGCTCCGAGTTTCCCGG 25

RESULT 35

AI432930

LOCUS

DEFINITION AI432930 28 bp mRNA linear EST 13-APR-1999

ch4c11.x1 NCI CGAP Lym12 Homo sapiens cDNA clone IMAGE:2121140 3',

similar to SW:PRP2 HUMAN P02812 SALIVARY PROLINE-RICH PROTEIN

PRECUSOR ; contains element MSR1 repetitive element ;, mRNA

sequence.

ACCESSION AI432930

VERSION AI432930.1

KEYWORDS EST. GI:4285621

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 28)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

DNA Sequencing catalog #: 11547-015

Life Technologies by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality

Insert length: 482 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 1.

Location/Qualifiers

1..28

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2121140"

/tissue_type="lymphoma, follicular mixed small and large

cell"

/lab_host="DH10B"

/clone_lib="NCI CGAP Lym12"

/note="Organ: lymph node; Vector: pCMV-SPORT6; Site 1:

SalI; Site 2: NotI; Cloned unidirectionally. Primer:

Oligo dT. Average insert size 1.25 kb. Life Technologies

catalog #: 11547-015"

Db	1	GGGGAAATTTCCCCGGGGCCCCCCCC	25		
RESULT 36	AZ628058	33 bp	DNA	linear	GSS 13-DEC-2000
LOCUS	1M0476122F	Mouse 10kb	plasmid	UUC1M	library Mus musculus genomic
DEFINITION	clone UUCG1M0476122 F, genomic survey sequence.				
ACCESSION	AZ628058				
VERSION	AZ628058.1	GI:11750248			
KEYWORDS	GSS.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 33)				
TITLE	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Ismail,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.				
JOURNAL	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts				
COMMENT	Unpublished (2000) Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SL/C, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0476 row: I column: 22 Seq primer: CGTTGTAACGACGCCAGT Class: plasmid ends High quality sequence stop: 33.				
FEATURES	Location/Qualifiers				
source	1..33				
	/organism="Mus musculus"				
	/mol_type="Genomic DNA"				
	/strain="C57BL/6J"				
	/db_xref="taxon:10090"				
	/clone="UUCG1M0476122"				
	/sex="Male"				
	/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"				
	/clone_lib="Mouse 10kb plasmid UUC1M library"				
	/note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi 4732114 gb AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."				
ORIGIN					
Query Match	36.3%;	Score 13.8;	DB 9;	Length 33;	
Best Local Similarity	72.0%;	Pred. No. 9e+05;			
Matches	18;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;

ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 41)
Humphray, S.J., Huckle, E. and Hunt, S.E.
Direct Submission
Submitted (14-NOV-2002) The Sanger Institute, Wellcome Trust Genome
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
humquerry@sanger.ac.uk Unpublished
This sequence was generated from the SP6 end of BAC 48J2. 48J2 is
part of the Daniokey BAC Library created by R. Plasterk and N.V.
Keygene.
Further details: http://www.sanger.ac.uk/projects/d_rerio/.

FEATURES

source

1..41

/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKY-48J2"
/tissue_type="Testis"
/note="vector pIndigoBAC-536"

ORIGIN

Query Match 36.3%; Score 13.8; DB 11; Length 41;
Best Local Similarity 72.0%; Pred. No. 9.3e+05;
Matches 18; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 11 CGCTGGGACTTCCAGGGGACTT 35
Db 40 CGCTGGGCTTTCCTCGGGTCTCTT 16

RESULT 39

AI520679/c

LOCUS ti06c03.x1 NCI CGAP Pan1 Homo sapiens cDNA clone IMAGE:2129668 3,
similar to TR:000599 O00599 CON1. ;, mRNA sequence.
AI520679
AI520679.1 GI:4434814

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 43)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 1286 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
Location/Qualifiers

FEATURES

source

1..43

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2129668"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI CGAP Pan1"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;

ORIGIN

Query Match 36.3%; Score 13.8; DB 1; Length 43;
Best Local Similarity 88.2%; Pred. No. 9.4e+05;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 15 GGGGACTTTCAGGGGG 31
Db 18 GGGGCTTTCCTCGGGG 2

RESULT 40

AI223998/c

LOCUS

DEFINITION

QX12H07.x1 NCI CGAP Lym12 Homo sapiens cDNA clone IMAGE:2001181 3,
similar to TR:004154 Q04154 SALIVARY PROLINE-RICH PROTEIN RP15
PRECURSOR. ;contains element MER22 repetitive element ;, mRNA
sequence.
AI223998
AI223998.1 GI:3806711

VERSION

KEYWORDS

SOURCE

ORGANISM

EST.

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 46)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
unknown library type
Trace considered overall poor quality
Insert Length: 1534 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
Location/Qualifiers

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

source

1..46

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2001181"

/tissue_type="lymphoma, follicular mixed small and large
cell"

/lab_host="DH10B"

/clone_lib="NCI CGAP Lym12"

/note="Organ: lymph node; Vector: pCMV-SPORT6; Site_1:
SalI; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dt. Average insert size 1.25 kb. Life Technologies
catalog #: 11547-015"

ORIGIN

Query Match 36.3%; Score 13.8; DB 1; Length 46;
Best Local Similarity 63.6%; Pred. No. 9.5e+05;
Matches 21; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 6 CTTTCGGCTGGGACTTTCAGGGGACTTTC 38
Db 38 CTTTTCCTGGGCTTTTTCGGGGGCTTTC 6

RESULT 41

AI544460/c

LOCUS

DEFINITION

fb75b10.x1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone
IMAGE:3717691 3, similar to SW:CA19 HUMAN P20849 COLLAGEN ALPHA
1(I)X CHAIN PRECURSOR. ;, mRNA sequence.
AI544460

ACCESSION

AI544460

```

VERSION      AI544460.1  GI:4461833
KEYWORDS     EST.
SOURCE       Danio rerio (zebrafish)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
              Cypriniformes; Cyprinidae; Danio.
REFERENCE    1 (bases 1 to 46)
AUTHORS      Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,P., Marra,M.,
              Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
              Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
              Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
              Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
              Waterston,R. and Wilson,R.
              WashU Zebrafish EST Project 1998
              Unpublished (1998)
              Contact: Stephen L. Johnson
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: zbrafish@watson.wustl.edu
              cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
              Matthew Clark. DNA Sequencing by: Washington University Genome
              Sequencing Center. Clone distribution: Genome Systems, St. Louis,
              Missouri (web address: www.genomesystems.com) (email contact:
              info@genomesystems.com) and Research Genetics, Huntsville, Alabama
              (web address: www.resgen.com) (email contact: info@resgen.com) and
              RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
              www.rzpd.de)
              Trace considered overall poor quality
              Possible reversed clone: similarity on wrong strand
              Seq primer: T7 Et from Amersham
              High quality sequence stop: 1
              POLYA=No.

FEATURES             Location/Qualifiers
     source           1..46
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="IMAGE:3717691"
                     /sex="mixed"
                     /tissue types="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="X11-blue MRP"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /notes="vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGGAGCGCGCCCTTTTCTTTTCTTTT3'];
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries. Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."

ORIGIN
Query Match      36.3%; Score 13.8; DB 1; Length 46;
Best Local Similarity 63.6%; Pred. No. 9.5e+05;
Matches 21; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 1 GGGGACTTTCGGCTGGGACTTTCACGGGGAC 33
    |||||
Db 34 GGGAAACCCCGAACCGGAATCCCGGGGCC 2
    |||||

```

```

RESULT 42
LOCUS       BJ057570
DEFINITION BJ057570 NIBB Mochii normalized Xenopus tailbud library Xenopus
              laevis cDNA clone XL104h07 5', mRNA sequence.
ACCESSION   BJ057570
VERSION     BJ057570.1  GI:17479651
KEYWORDS    Xenopus laevis (African clawed frog)
SOURCE      Xenopus laevis
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
              Xenopodinae; Xenopus; Xenopus.
REFERENCE    1 (bases 1 to 46)
AUTHORS      Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
              Kohara,Y.
              Expressed genes in X. laevis embryo
              Unpublished (2001)
              Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshini@genes.nig.ac.jp
              The information of this clone is available through the following
              URL.
              http://xenopus.nibb.ac.jp.

FEATURES             Location/Qualifiers
     source           1..46
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="XL104h07"
                     /tissue type="whole embryo"
                     /dev_stage="stage 25"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"

ORIGIN
Query Match      36.3%; Score 13.8; DB 3; Length 46;
Best Local Similarity 66.7%; Pred. No. 9.5e+05;
Matches 18; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 12 GCTGGGACTTTCACGGGGACTTTC 38
    |||||
Db 14 GCTGGGAAGTTGACGGGNGTTTANCC 40
    |||||

RESULT 43
LOCUS       AU102496
DEFINITION AU102496 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
              HEP22967, mRNA sequence.
ACCESSION   AU102496
VERSION     AU102496.1  GI:13552016
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Homnidae; Homo.
REFERENCE    1 (bases 1 to 50)
AUTHORS      Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J.,
              Hata,H., Ota,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K.,
              Sakaki,Y., Nakamura,Y., Suyama,A. and Sugano,S.
              Diverse transcriptional initiation revealed by fine, large-scale
              mapping of mRNA start sites
              EMBO Rep. 2 (5), 388-393 (2001)
              11375929
              Contact: Yutaka Suzuki

```

Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: ysuzuki@ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and
Sugano,S. Construction and characterization of a full
length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2),
149-156 (1997).

FEATURES

source
1..50
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HEP22967"
/clone_lib="Sugano Homo sapiens cDNA library"

ORIGIN

Query Match 36.3%; Score 13.8; DB 1; Length 50;
Best Local Similarity 63.6%; Pred. No. 9.6e+05;
Matches 21; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 6 CTTTCGGCTGGGACTTTCAGGGGGACTTCC 38
|||||
DB 2 CTTTCGGCTGGGCTGTTTCTCCGACGGTTTC 34
|||||

RESULT 44

AU107397/c
LOCUS AU107397 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
LNG13032, mRNA sequence.
ACCESSION AU107397
VERSION AU107397.1 GI:13556918
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM

REFERENCE

AUTHORS Suzuki,Y., Taira,H., Tsunoda,T., Mizushima-Sugano,J., Sese,J.,
Hata,H., Oca,T., Isogai,T., Tanaka,T., Morishita,S., Okubo,K.,
Sakaki,Y., Nakamura,Y., Suyama,A. and Sugano,S.
TITLE Diverse transcriptional initiation revealed by fine, large-scale
mapping of mRNA start sites
JOURNAL EMBO Rep. 2 (5), 388-393 (2001)
PubMed 11375929
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: ysuzuki@ims.u-tokyo.ac.jp

Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and
Sugano,S. Construction and characterization of a full
length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2),
149-156 (1997).

FEATURES

source
1..50
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="LNG13032"
/clone_lib="Sugano Homo sapiens cDNA library"

ORIGIN

Query Match 36.3%; Score 13.8; DB 1; Length 50;
Best Local Similarity 63.6%; Pred. No. 9.6e+05;
Matches 21; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 GGGGACTTTCGGCTGGGACTTTCAGGGGGAC 33
|||||
DB 49 GGAGACAATCTCTTGTCTGACTGTCCAGGCATC 17
|||||

RESULT 45

CR178464/c
LOCUS CR178464 50 bp DNA linear GSS 06-JUL-2004
DEFINITION Forward strand read from insert in 3'HPRT insertion targeting and
chromosome engineering clone MHP340e12, genomic survey sequence.
ACCESSION CR178464
VERSION CR178464.1 GI:49957313
KEYWORDS GSS; genome survey sequence; MICER.
SOURCE Mus musculus (house mouse)
ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 50)
AUTHORS Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
Rogers,J. and Bradley,A.

TITLE

Direct Submission
Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. http://www.sanger.ac.uk/MICER

FEATURES

source
1..50
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHP340e12"
/clone_lib="MHPp"

ORIGIN

Query Match 36.3%; Score 13.8; DB 11; Length 50;
Best Local Similarity 88.2%; Pred. No. 9.6e+05;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 GCTGGGACTTCCAGG 28
|||||

DB 38 GCTGGCGACATTCCAGG 22
|||||

RESULT 46

BH812215
LOCUS BH812215 32 bp DNA linear GSS 02-MAY-2002
DEFINITION SALUK_061414 Arabidopsis thaliana TDNA insertion lines Arabidopsis
thaliana genomic clone SALK_061414, genomic survey sequence.
ACCESSION BH812215
VERSION BH812215.1 GI:20390670
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE

AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shinn,P., Zimmerman,J. and Ecker,J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
JOURNAL Unpublished (2001)
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: eckersalk.edu
This is single pass sequence recovered from the left border of
TDNA.

Class: TDNA tagged.
Location/Qualifiers
1..32
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"

FEATURES

source
1..32
Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"

/ecotype="Col-0"
 /db_xref="taxon:3702"
 /clone_lib="SALK_06414"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /notes="PCR was performed on Arabidopsis thaliana lines
 each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN

Query Match 35.8%; Score 13.6; DB 9; Length 32;
 Best Local Similarity 80.0%; Pred. No. 1.1e+06;
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 ACTTTCCGTTGGGACTTTC 24
 ||||| |||||
 Db 9 ACTTTCCGTTAGGACTTTC 28

RESULT 47

AZ775335 40 bp DNA linear GSS 16-FEB-2001
 LOCUS 2M007G07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 DEFINITION clone UUGC2M0007G07 R, genomic survey sequence.

ACCESSION AZ775335
 VERSION
 KEYWORDS
 SOURCE GSS.

ORGANISM Mus musculus (house mouse)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridea; Muridae; Murinae; Mus.

1 (bases 1 to 40)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
 Reilly,M., Rose,R., Stokes,R., Tingey,A., von
 Niederhausern,A. and Wright,D. Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0007 row: G column: 07

Seq primer: CACACAGGAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 40.

Location/Qualifiers

FEATURES

source

1..40
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0007G07"
 /sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /note="Vector: FWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

ORIGIN

Query Match 35.8%; Score 13.6; DB 9; Length 40;
 Best Local Similarity 67.9%; Pred. No. 1.1e+06;
 Matches 19; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 GGGGACTTTCCGCTGGGACTTTCCAGG 28
 ||||| ||||| ||||| |||||
 Db 7 GGGGGCTTGCAGTTGGGTTCTGCGCTGG 34

RESULT 48

TA26H11Q 41 bp DNA linear GSS 13-DEC-2000
 LOCUS T. brucei sheared genomic DNA clone 26h11, reverse sequence,
 DEFINITION genomic survey sequence.

ACCESSION AL452967
 VERSION AL452967.1 GI:11850952
 KEYWORDS GSS.

SOURCE Trypanosoma brucei

ORGANISM Trypanosoma brucei

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

1 (bases 1 to 41)

Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,

Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,

Melville,S.E., Rajandream,M.A. and Barrell,B.G.

Direct Submission

Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
 project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
 Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
 nh1@sanger.ac.uk

Constructed at the Institute for Genomic Research (TIGR),
 Rockville, MD. Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
 to give a tight size distribution (
 4 kb). The v + i method used for the library construction is
 described in detail in Smith, H. and Venter, J.C. (Making small
 insert libraries for whole genome shotgun sequencing projects. In
 Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org

Details of T. brucei sequencing at the Sanger Centre are available
 at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

source

1..41
 /organism="Trypanosoma brucei"
 /mol_type="genomic DNA"
 /strain="TREU927"
 /db_xref="taxon:5691"
 /clone="26h11"

ORIGIN

Query Match 35.8%; Score 13.6; DB 11; Length 41;
 Best Local Similarity 67.9%; Pred. No. 1.1e+06;
 Matches 19; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 GACTTTCCGCTGGGACTTTCCAGGGG 31
 ||||| ||||| ||||| |||||
 Db 13 GAGTGCTGCAGCGCATTTGCCAGGGG 40

```

RESULT 49
W10989/c
LOCUS
DEFINITION
ma41b12.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:313247 5', mRNA sequence.

ACCESSION
W10989
VERSION
W10989.1 GI:1285294
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 45)
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE
The WashU-HMI Mouse EST Project
JOURNAL
Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI: 203863
Seq primer: -21M13
High quality sequence stop: 1.
FEATURES
Location/Qualifiers
1..45
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:313247"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares mouse p3NMF19.5"
/vector="Vector: pT7T3D (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCATCTGAGTCGGAGCGCCGCAATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
```

ORIGIN

```

Query Match      35.8%; Score 13.6; DB 8; Length 45;
Best Local Similarity 67.9%; Pred. No. 1.1e+06;
Matches 19; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 8 TTCCGCTGGGACTTTCACGGGGGACTT 35
   ||| ||| ||| ||| ||| ||| |||
Db 28 TTGAGCTTGAACGTGTCACCTGGAACCT 1
```

```

RESULT 50
A1635210/c
LOCUS
DEFINITION
t222e06.x1 NCI CGAP Ut2 Homo sapiens cDNA clone IMAGE:2289346 3'
similar to SW:EXTN_DAUCA P06599 EXTENSIN PRECURSOR. ;, mRNA
sequence.
ACCESSION
A1635210
VERSION
A1635210.1 GI:4686540
KEYWORDS
EST.
```

```

SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 46)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 1695 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
FEATURES
Location/Qualifiers
1..46
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2289346"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/clone_lib="NCI CGAP Ut2"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"

ORIGIN
Query Match      35.8%; Score 13.6; DB 1; Length 46;
Best Local Similarity 61.1%; Pred. No. 1.1e+06;
Matches 22; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1 GGGGACTTTCGGCTGGGACTTTCAGGGGACTTT 36
   ||| ||| ||| ||| ||| ||| |||
Db 41 GGGGTTTTTTCGGTGGGGAAACCTTGGAAACCTT 6

Search completed: February 15, 2006, 21:10:37
Job time : 2541.47 secs
```

THIS PAGE BLANK (USPTO)